

Stefan Engelberg, Peter Meyer (Mannheim)

Das *Lehnwortportal* Deutsch als kontaktlinguistisches Forschungsinstrument

Abstract: The web portal *Lehnwortportal Deutsch* <lwp.ids-mannheim.de>, developed at the Institute for the German Language (IDS), aims to provide unified access to a growing number of lexicographical resources on German loanwords in other languages. This paper discusses different possibilities of creating an onomasiological access structure for portal users. We critically examine the meaning list of the “World Loanword Database” project (Haspelmath/Tadmor 2009a) as well as WordNet-based taxonomies and propose a new way of inductively creating a semantic classification scheme that takes both hyperonymic relations and semantic fields into account. We show how such a classification can be integrated into the underlying graph-based data representation of the *Lehnwortportal* and thus be exploited for advanced onomasiological search options.

1. Das *Lehnwortportal* Deutsch

Die am Institut für Deutsche Sprache entwickelte und gepflegte, frei zugängliche Online-Anwendung „*Lehnwortportal Deutsch*“¹ ermöglicht Nutzern einen einheitlichen Zugriff auf eine wachsende Anzahl von Wörterbüchern zu deutschen Lehnwörtern in anderen Sprachen. Damit fokussiert das Portal – anders als in der Sprachkontaktlexikographie üblich – nicht auf die Zielsprache der Entlehnungsprozesse, sondern auf die Quellsprache. Jedes in einem beliebigen Lehnwörterbuch des Portals gebuchte deutsche Etymon (das häufig einer älteren Sprachstufe oder einem Dialekt angehört) wird manuell einem deutschen Portal-Metalemma, das von diasystematischer Variation abstrahiert, zugeordnet. Abgesehen von den üblichen, insbesondere lemmabasierten, makrostrukturellen Zugriffsmöglichkeiten auf die Einzelwörterbücher werden durch diese portalweite lexikographische Vernetzung zum einen komplexe wörterbuchübergreifende Suchfunktionen möglich (vgl. Meyer 2013a). Zum anderen fungieren die – auch untereinander aufgrund von Derivations- und Kompositionsbeziehungen verknüpften – Metalemmata als Lemmata eines „umgekehrten Lehnwörterbuchs“ (vgl. Engelberg 2010, Meyer/Engelberg 2011), also eines automatisch erzeugten zusätzlichen Portalwörterbuchs der deutschen Herkunftswörter, dessen Einträge

1 Vgl. <<http://lwp.ids-mannheim.de/>> [aufgerufen am 04.07.2015].

die Entlehnungswege deutscher Etyma in die verschiedenen Nehmersprachen der integrierten Lehnwörterbücher nachvollziehbar machen.

Eine der zentralen Aufgaben beim strukturellen Ausbau des Portals ist die Schaffung eines onomasiologischen Zugangs, der es erlauben soll, Abfragen über Sprachen, Zeiten und formale Eigenschaften mit solchen über semantisch-diskursive Eigenschaften von Herkunfts- und Lehnwörtern zu verknüpfen. In diesem Beitrag sollen die Vor- und Nachteile verschiedener onomasiologischer Konzeptionen und deren Umsetzung in eine onomasiologische Zugriffsstruktur diskutiert werden.

2. Onomasiologische Strukturierungsoptionen

2.1. Anforderungen an ein onomasiologisches System

Unter einem onomasiologischen System soll hier ein Klassifikationssystem verstanden werden, das Wortbedeutungen² nach semantischen, pragmatischen und/oder diskursiven Eigenschaften für lexikographische Zwecke klassifiziert. Die Überlegungen dazu bewegen sich damit im Umfeld von Konzepten wie Sinnbezirken, Ontologien, Sachgruppen, Sachbereichen, lexikalischen Netzen und Thesauri. Ziel der hier formulierten Überlegungen ist es primär, verschiedene Optionen für solche Systeme hinsichtlich ihrer lexikographischen Praktikabilität zu sichten. Grundlegende semantische Überlegungen müssen weitgehend außen vor bleiben.

Über seine Nützlichkeit für die Bereitstellung entsprechender Suchfunktionen im Lehnwortportal hinaus sollte das verwendete onomasiologische System im Idealfall die folgenden Anforderungen erfüllen:

1. Es sollte in seinen Konzepten und Relationen semantisch gut begründet und definiert sein.
2. Es sollte in der Annotation möglichst einfach und eindeutig anwendbar sein.
3. Es sollte für Lehnwörterbücher verschiedener Typen und Zielsprachen verwendbar sein.

Bedingung 1 verlangt zum einen, dass die einzelnen Klassen hinreichend präzise definiert sind, zum anderen aber auch, dass die grundlegenden semantisch-diskursiven Konzepte – seien es nun hyperonymische Relationen, Sinnbezirke oder

2 Um die Redeweise von einem onomasiologischen System zu rechtfertigen, das ja einen Weg von lexikalischen Inhalten zu lexikalischen Formen nahelegt, müssen Wortbedeutungen hier genauer als disambiguierte Lexeme verstanden werden.

Diskursbereiche – hinreichend geklärt sind. Bedingung 2 ist dadurch operationalisierbar, dass eine manuelle semantische Annotation möglichst wenig Zeit erfordert und gleichzeitig mit einer hohen Annotatorenübereinstimmung einhergeht. Bedingung 3 schließlich hebt erstens darauf ab, dass Lehnwörterbücher bezüglich des lexikalischen Bereichs, den sie erfassen, oft bestimmte Fokussierungen oder Einschränkungen aufweisen, etwa hinsichtlich der Aufnahme von Internationalismen, fachsprachlich-wissenschaftlicher Lexik oder sogar Eigennamen. Zweitens unterscheiden sich die Kontaktsituationen, in denen einzelne Sprachen aus dem Deutschen entlehnt haben, erheblich, und damit auch der entlehnte Wortschatz. So lassen sich unterscheiden: (i) langandauernder Kontakt an Bevölkerungsgrenzen; Entlehnungen z. B. ins Slowenische, Polnische, Niederländische, Sorbische; (ii) Kontakt durch Emigration mit Sprachinselbildung; Entlehnungen z. B. ins Rumänische; (iii) Kontakt durch Elitenaustausch; Entlehnungen z. B. ins Japanische, Russische, Englische; (iv) Kontakt in kolonialen oder missionarischen Kontexten; Entlehnungen z. B. ins Basaa (Kamerun), Samoanische, Labrador-Inuit, Yapesische (Mikronesien) oder Kuanua (Papua-Neuguinea). Dazu kommen indirekte Entlehnungen über Mittlersprachen, also Entlehnungen ohne direkten Kontakt mit dem Deutschen, so etwa über die Entlehnungskette Polnisch > Weißrussisch > Russisch ins Usbekische und andere Sprachen Zentralasiens oder über das Niederländische in die Sprachen Südasiens und in Niederländisch basierte Kreolsprachen.

Über die drei genannten Bedingungen hinaus sollte das onomasiologische System natürlich möglichst große Teile des verzeichneten Wortschatzes im Portal abdecken. Es ist allerdings nicht erforderlich, dass der Lemmabestand vollständig semantisch-thematisch erschlossen wird. Schließlich geht es nicht darum, die exhaustive Anwendbarkeit eines onomasiologischen Systems auf den Wortschatz einer Sprache nachzuweisen, sondern einen onomasiologischen Zugriff zu entwickeln, der lehnwortlexikalisch relevante Abfragen über semantisch-thematische Bereiche erlauben soll. Dabei ist auch zu berücksichtigen, dass mit Sachgruppenwörterbüchern wie etwa dem Dornseiff (2004), in denen onomasiologische Systeme und Zugriffe ihr lexikographisches Hauptanwendungsgebiet finden, ganz andere Nutzungsszenarien verbunden sind als mit onomasiologischen Zugriffen in Lehnwörterbüchern. Während jene vor allem die Ausdrucksfindung bei der Sprachproduktion befördern möchten, sollen letztere die Beantwortung von sprachhistorischen, sprachkontakttheoretischen oder kulturgeschichtlichen Fragen unterstützen. Eine Übernahme der entsprechenden Systeme aus der Sachgruppenlexikographie ist also nicht unbedingt naheliegend und ist im Rahmen des hier beschriebenen Projekts auch (noch) nicht verfolgt worden.

Im Folgenden sollen drei Optionen für die ontologische Strukturierung des Wortbestands des Portals diskutiert werden: (i) eine Klassifizierung auf der Basis der Bedeutungsklassen der Loanword Typology Meaning List des Leipziger Lehnwortprojekts, (ii) eine Anbindung an das Netz semantischer Relationen in GermaNet und (iii) der Versuch einer induktiven Entwicklung eines onomasiologischen Systems aus den Lehnwortdaten. Im Zuge dieser Diskussion werden wir Daten aus folgenden Wörterbüchern diskutieren:

- *W-Pol*: „Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache“ (de Vincenz/Hentschel/Brüggemann 2010; im Lehnwortportal enthalten); verzeichnet keine Lehnwörter, für die das Deutsche lediglich als Mittlersprache diene.
- *W-Frz*: „Das deutsche Lehngut im Französischen als Zeugnis für den Wissenstransfer im 20. Jahrhundert“ (Sarcher 2001); starker Anteil an wissenschaftlichem Wortschatz aus dem 20. Jh.
- *W-Slv*: „Deutsche Lehnwörter im Slovenischen“ (Striedter-Temps 1963; im Lehnwortportal enthalten); breite semantische Entlehnungsbereiche über einen langen Entlehnungszeitraum.
- *W-Pzf*: „Wörterbuch deutscher Lehnwörter in den Sprachen des Südpazifiks“ (in Vorbereitung am Institut für deutsche Sprache); Lehnwörter größtenteils aus der deutschen Kolonialzeit (v. a. in den Bereichen Schule, Religion, Verwaltung, Alltagsgegenstände).

Die folgenden Bemerkungen repräsentieren den gegenwärtigen Stand der noch nicht abgeschlossenen konzeptionellen Überlegungen zu einem onomasiologischen Zugriff für das Lehnwortportal.³

2.2. Bedeutungsklassen der „Loanword Typology Meaning List“

Die „World Loanword Database“ (Haspelmath/Tadmor 2009a, 2009b) basiert auf der sogenannten „Loanword Typology Meaning List“ (LTML), einer Liste von 1460 Konzepten, die 24 semantischen Klassen zugeordnet werden. Für diese Konzepte wurde zu 41 Zielsprachen ermittelt, ob und aus welcher Sprache die sie ausdrückenden Lexeme entlehnt wurden. Die 24 semantischen Klassen sind die folgenden:

3 Aus expositorischen Gründen konzentrieren wir uns im Folgenden auf die Klassifizierung von Lehnwortbedeutungen; dies schließt selbstverständlich eine analoge onomasiologische Behandlung von Herkunftswörtern nicht aus.

01: THE PHYSICAL WORLD	09: BASIC ACTIONS AND TECHNOLOGY	17: COGNITION
02: KINSHIP	10: MOTION	18: SPEECH AND LANGUAGE
03: ANIMALS	11: POSSESSION	19: SOCIAL AND POLITICAL RELATIONS
04: THE BODY	12: SPATIAL RELATIONS	20: WARFARE AND HUNTING
05: FOOD AND DRINK	13: QUANTITY	21: LAW
06: CLOTHING AND GROOMING	14: TIME	22: RELIGION AND BELIEF
07: THE HOUSE	15: SENSE PERCEPTION	23: MODERN WORLD
08: AGRICULTURE AND VEGETATION	16: EMOTIONS AND VALUES	24: MISCELLANEOUS FUNCTION WORDS

Wir haben diese Klassifikation auf 720 Lehnwörter aus 30 Sprachen aus dem Südpazifik-Wörterbuch *W-Pzf* angewendet. Dabei tauchten zwei hauptsächliche Probleme auf. Das erste Problem betrifft die geringe Differenziertheit der Klassifikation. Die Lehnwörter verteilen sich erwartungsgemäß sehr ungleich auf die 24 Klassen, zwischen 0 (KINSHIP) und 105 (RELIGION AND BELIEF). Für die stark frequentierten Klassen läge hier eine größere Differenzierung nahe. Als problematisch erweist sich hier vor allem die Fokussierung der LTML auf prämoderne Konzepte, die zu einer geringen Differenzierung in Bereichen des modernen Lebens führen (Technik, Wissenschaft, Gesellschaft). Dies wirkt sich natürlich auch erheblich in Lehnwörterbüchern wie *W-Frz* aus, das nur Entlehnungen aus dem 20. Jh. dokumentiert.

Während dieses Problem gegebenenfalls durch eine eigene Subklassifikation behoben werden könnte, ist das zweite Problem substantiellerer Natur. Ein kohärenter Handlungs- und Diskursbereich, der Entlehnungen aus dem Deutschen in deutsch-kolonialen Zusammenhängen befördert hat, ist die Schule. Z. B. sind hier *Bleistift*, *malen*, *Griffel*, *Schule*, *Papier*, *auswendig*, *Tafel* und viele andere deutsche Etyma zur Quelle von Entlehnungen geworden. Ein Feld SCHULE kommt in der Leipziger Klassifikation aber nicht vor und ließe sich auch nicht als Subfeld eines schon bestehenden Feldes konstruieren. Vielmehr fallen Lehnwörter aus dem Schulkontext in verschiedene Felder der LTML: 17 COGNITION (hier wären *Schule* und *auswendig* einzuordnen) / 18 SPEECH AND LANGUAGE (*Papier*, *Bleistift*, *Buch*) / 23 MODERN WORLD (*Brief*).

Insgesamt ließen sich zwei Drittel der Lehnwörter in der LTML-Klassifikation gut, wenn eben auch nicht besonders differenziert, verorten. Für das andere

Drittels mussten weitere Klassen wie etwa SCHOOL, TRANSPORT oder ADMINISTRATION hinzugefügt werden, um zu einer Klassifikation zu gelangen, die die für die spezifische Kontaktsituation relevanten Bereiche angemessen differenziert berücksichtigt und semantisch-diskursiv zusammengehörige Lehnwortgruppen nicht klassifikatorisch auseinanderreißt. Letztlich zeigt sich aber, dass aus der Geeignetheit der Feldgliederung der „Loanword Typology Meaning List“ für typologische Fragestellungen nicht die Geeignetheit für konkrete historisch und sprachlich spezifische Lehnwortschätze folgt.

2.3. Anbindung an GermaNet

Während die onomasiologische Ordnung der LTML auf der Zuordnung von Wortbedeutungen zu semantischen Feldern basiert, vernetzen lexikalisch-semantische Wortnetze wie das englischsprachige WordNet (Fellbaum 1998) und das deutschsprachige GermaNet (Hamp/Feldweg 1997, Henrich/Hinrichs 2010) den Wortschatz auf der Basis semantischer Relationen zwischen Wörtern, wobei in unserem Zusammenhang vor allem die das Wortnetz dominierenden hyperonymischen Relationen von Interesse sind.

Die Probleme bei den Versuchen, Lehnwortbedeutungen an die hyperonymische Struktur eines solchen Wortnetzes anzubinden, sollen im Folgenden am Beispiel des polnischen Lehnwörterbuchs *W-Pol* erläutert werden. Hier stellt sich zunächst die Frage nach der Auswahl einer geeigneten WordNet-Ressource. Vor dem Hintergrund, dass später Abfragen über unterschiedliche Wörterbücher und Nehmersprachen hinweg möglich sein sollen, ist die Verwendung je verschiedener einzelsprachlicher Wortnetze nicht sinnvoll. Der Verwendung einer Ressource wie EuroWordNet (Vossen et al. 1998), die über einen „interlingualen Index“ je eigenständige einzelsprachliche Wortnetze aufeinander abbildet, steht unter anderem die kleine Anzahl der in diesem mittlerweile abgeschlossenen Projekt berücksichtigten Sprachen entgegen, aber natürlich auch die grundsätzliche Problematik der Verknüpfung von womöglich sehr unterschiedlich strukturierten Wortnetzen. Unter rein praktischen Gesichtspunkten ist für das Lehnwortportal Deutsch die sprachübergreifende Zuordnung von Wortbedeutungen zu „passenden“ GermaNet-Synsets vermutlich die einzige gangbare, wenngleich theoretisch nicht ganz befriedigende Vorgehensweise.

Es erweist sich dabei, dass aus verschiedenen Gründen Lehnwortbedeutungen oft mehrere Hyperonyme in GermaNet zugeordnet werden müssen. Das ist insofern ein Problem, als die Zuordnung eines Wortes zu zwei Hyperonymen außer in Fällen, in denen es in GermaNet Mehrfachvererbung gibt, auch zwei wohlunterschiedene Bedeutungspositionen des Wortes erwarten lässt. Diese Voraussetzung

ist aber nicht immer gegeben. Zum einen verzichten Wörterbücher häufig auf Lesartendifferenzierungen, die für die Zuordnung zu GermaNet-Hyperonymen erforderlich sind. So wird *rajsplata* in *W-Pol* nur in einer Bedeutungsposition dargestellt, die mit der Beschreibung 'Arbeitstisch, gußeiserne, glatte, waagerechte Tischplatte' versehen ist. Das bedingt die doppelte Einbindung in GermaNet als |>*Tisch*2 >*Möbel*1 >*Artefakt*1| und als |>*Tischplatte*1 >*Platte*5 >*Brett*2 >*Artefakt*1>|. ⁴ Mehrfache Einordnungen in die GermaNet-Hierarchie treten des Weiteren auch bei systematischer Polysemie auf, soweit diese nicht im Wörterbuch abgebildet ist: *jarmuż* 'Grünkohl' ist sowohl als |>*Grünkohl*1 >... >*Pflanze*1| wie als |>*Grünkohl*2 >... >*Gemüse*1 >... >*Nahrung*1> ...| zu klassifizieren. Drittens treten Mehrfachzuordnungen bei unklaren semantischen Differenzierungen in GermaNet auf. So könnte *śrubsztak*, mit der Bedeutungsbeschreibung 'Vorrichtung zum Festhalten von zu bearbeitenden Gegenständen, Schraubstock' je nach Interpretation der GermaNet-Hyperonyme als |>*Spannvorrichtung*1 >*Vorrichtung*1> ...|, als |>*Haltevorrichtung*1 >*Vorrichtung*1> ...| oder als |>*Spezialwerkzeug*1 >*Werkzeug*1> ...| eingeordnet werden.

Ein zweiter Problemkomplex bei der Verwendung von GermaNet entsteht dadurch, dass die Hierarchisierungen in GermaNet zum Teil den semantischen Differenzierungen im Lehnwortbereich nicht angemessen sind. So ist bei Verben oft nur eine sehr grobe Einordnung in GermaNet möglich: *takielować*, in der Bedeutung 'die Takelage eines Schiffes vorbereiten' ließe sich sowohl unter |>*einrichten*3|, |>*vorbereiten*1| oder |>*fertig machen*1| einordnen, während es sehr wohl ein substantivisches Synset *Takelage*1 gibt. Dabei ist zudem nicht klar, welche Benutzerinteressen Abfragen befriedigen könnten, die auf solchen Zuordnungen basieren. Neben mangelnder Spezifik in einigen Bereichen ist das Wortnetz in anderen Teilen wiederum sehr filigran strukturiert. Das macht die Einordnung von Fachwortschatz, die ja lediglich aufgrund der in den Wörterbüchern gegebenen Paraphrasen erfolgen kann, oft schwierig. Ein weiteres Problem ist die Behandlung von prämodernem Wortschatz, der, ebenso wie fachsprachliche Bereiche, unter Umständen aufwändige Erweiterungen des Wortnetzes erforderlich machen kann.

Die vorgetragenen Bedenken legen nahe, Lehnwortbedeutungen nicht direkt semantisch passenden Synsets in GermaNet oder auch nur möglichst weit unten in der Hierarchie angeordneten Hyperonymen zuzuordnen. Auch für Benutzer des Lehnwortportals wäre es im Einzelfall eine anspruchsvolle Aufgabe, ein zur

4 Die Zahlen indizieren Bedeutungspositionen der Hyperonyme in GermaNet. Die Beispiele beziehen sich auf Version 7.0 von GermaNet.

Problemstellung gut passendes Synset in den verzweigten Hierarchien eines Wortnetzes überhaupt ausfindig zu machen. Eine alternative Vorgehensweise besteht also darin, die Wortnetzhierarchie in einer bestimmten Tiefe abzuschneiden und Lehnwortbedeutungen auf eine begrenzte Anzahl von Synsets einer mittleren oder oberen Hierarchieebene abzubilden. Solche Synsets müssten auch in irgendeinem für die Sprachkontaktforschung relevanten Sinne besonders grundlegende semantische Konzepte repräsentieren. Die zunächst vielleicht naheliegende Verwendung nur der hierarchisch höchsten Konzepte (entsprechend etwa den *unique beginners* oder *supersenses* im Princeton-WordNet) scheitert jedoch letztlich an deren zu hohem Abstraktionsgrad, ähnlich wie Versuche, eine formale und daher für maschinelles logisches Schlussfolgern ausgelegte Top-Level-Ontologie wie SUMO (Niles/Pease 2001) auf eine WordNet-artige Ressource abzubilden, wie das für das Princeton-WordNet durchgeführt worden ist (Niles/Pease 2003).

Vielsprechender sind Versuche, nicht (nur) die am deutlichsten allgemeinen, sondern die in einem näher zu definierenden Sinne für onomasiologische Zwecke am ehesten relevanten Synsets auszuwählen. Zwei Ansätze, die wir jedoch bislang nicht systematisch auf ihre Tragfähigkeit im vorliegenden Kontext geprüft haben, sind die folgenden:

- Im EuroWordNet-Projekt wurden wortnetzübergreifend etwa 1000 sogenannte Base Concepts definiert, die den für die semantischen Lexikalisierungsmuster der einzelnen Wortnetze zentralen Synsets entsprechen und dementsprechend möglichst zugleich hoch in der Netzhierarchie und reich an Relationen zu anderen Synsets sein sollen (Vossen et al. 1998). Davon kommen ca. 160 sogenannte Core Base Concepts in mindestens drei von den sieben Wortnetzen des Projekts (Niederländisch, Italienisch, Spanisch, Deutsch, Französisch, Tschechisch, Estnisch) vor.
- Linguistisch womöglich noch interessanter sind Versuche, unter den Synsets solche auszuwählen, die *basic-level concepts* im Sinne von Rosch (1977) entsprechen, also unter Umständen eine basalere kognitive Relevanz haben, da sie semantisch verhältnismäßig homogen und zugleich deutlich abgegrenzt von auf derselben Hierarchieebene benachbarten Konzepten sind. Die eben erwähnten Base Concepts sind in den meisten Fällen eine oder zwei Hierarchieebenen oberhalb von typischen basic-level concepts, liegen also eher auf der Ebene von *Möbel/Einrichtungsgegenstand* als auf der von *Tisch* oder *Spiegel*. Es gibt Versuche, mit automatisierten Verfahren eine eng begrenzbare Anzahl von Kandidaten für WordNet-Synsets auf dem basic level ausfindig zu machen (vgl. Izquierdo/Suárez/Rigau 2007, Lin 2010), wobei die Güte der Verfahren

typischerweise mit ihrer Eignung für maschinelle Sprachverarbeitungsaufgaben wie etwa Lesartendisambiguierung beurteilt wird.

In jedem Fall wird für die Zwecke der Einrichtung einer onomasiologischen Suche ein im weitesten Sinne induktives, an den jeweils betrachteten Einzelfällen orientiertes Vorgehen erforderlich sein, da der Umfang der hyperonymischen Klassen handhabbar und die Taxonomie selber den Besonderheiten der jeweiligen Lehnwortschätze angemessen sein muss.

2.4. Eine induktive Ontologie

Die beiden vorgestellten Lösungen zur Unterstützung onomasiologischer Zugänge sind von zwei unterschiedlichen semantischen Konzepten ausgegangen, semantischen Feldern einerseits und hyperonymischen Relationen andererseits. Beide Arten inhaltlicher Zuordnung können natürlich für eine lexikalische Einheit vorgenommen werden und sind als Grundlagen onomasiologischer Zugriffsstrukturen konzipierbar: So lässt sich *Straßenbahn* in einer Taxonomie $|\langle \text{Schienenfahrzeug} \langle \text{Fahrzeug} \langle \text{Artefakt} |$ hyperonymisch verankern ebenso wie es sich in einem semantisch-diskursiven Feld TRANSPORT & VERKEHR verorten lässt.

Angesichts der Probleme, die die Anwendung von LTML und GermaNet auf unseren Lehnwortbestand begleiteten, soll im Folgenden versucht werden, eine Klassifikation induktiv aus den spezifischen lexikalischen Daten der Lehnwörterbücher zu gewinnen. Klassifiziert werden sollen dabei disambiguierte Lexeme, also Wortbedeutungen, insofern sie in den integrierten Wörterbüchern als Bedeutungspositionen ausgezeichnet sind. Die Wortbedeutungen sollen sowohl bezüglich hyperonymischer Relationen als auch in thematische Felder eingeordnet werden. Das Wesen der Hyperonymie als semantischer Relation lässt sich logisch-semantisch über Implikationen (und strukturelle Zusatzannahmen) bestimmen und soll hier nicht weiter diskutiert werden. Weitgehend unbeantwortet und in der Literatur kaum diskutiert ist die Frage, welche Relation eigentlich zwischen einem Wort bzw. einer Wortbedeutung und einem semantischen Feld, thematischen Feld, Sinnbezirk etc. besteht. Wir konzipieren diese Beziehung hier als eine Relation zwischen einer Wortbedeutung W^B als semantischer Einheit und einem thematischen Feld T als einer textuell-pragmatischen Einheit. Ein thematisches Feld lässt sich dabei als ein thematisch gebundenes virtuelles Textkorpus verstehen und wäre damit ähnlich operationalisiert wie ein Diskurs in der Auffassung von Busse/Teubert (1994: 15). Die Relation zwischen W , B und T sei wie folgt definiert: Ein Wort W in der Bedeutung B gehört einem thematischen Feld T

dann an, wenn W^B typischerweise in T vorkommt. An diese Zuordnung sind die folgenden Erwartungen geknüpft:⁵

- W^B ist mit T stärker kognitiv assoziiert als mit den weitaus meisten anderen thematischen Feldern.⁶
- W^B ist stark mit vielen anderen T zugeordneten lexikalischen Einheiten kognitiv assoziiert.
- W^B tritt in T relativ zur jeweiligen T -Korpusgröße häufiger auf als in den weitaus meisten anderen thematischen Feldern.
- W^B kookkuriert überzufällig oft mit vielen anderen T zugeordneten lexikalischen Einheiten.⁷

Die doppelte Kodierung von Wortbedeutungen in Bezug auf hyperonymische Relationen einerseits und thematische andererseits erlaubt es nun, wichtige Differenzierungen vorzunehmen. Dies soll an einem Wortschatzausschnitt aus *W-Pzfillustriert* werden. In den Sprachen des Südpazifiks findet sich eine größere Anzahl von Lehnwörtern aus dem Tierbereich, die unter anderem auf folgende deutsche Etyma zurückgehen: *Biene, Büffel, Esel, Frosch, Gans, Hund, Kakerlake, Kalb, Kamel, Katze, Kuh, Lamm, Löwe, Ochs, Rindvieh, Ross, Schaf, Schlange, Taube, Wachtel* und *Ziege*. Aus der Perspektive semantischer Relationen lassen sich alle diese Wörter als Hyponyme von *Tier* beschreiben. Thematisch gehören die entsprechenden Lehnwörter aber ganz unterschiedlichen Bereichen an. Während einige der Wörter im Kontext des Lebensumfelds (HAUSTIERE & VIEHZUCHT) Verwendung finden (z. B. *Gans, Katze*), treten andere Wörter als Konsequenz von Bibelübersetzungen in die Sprachen des Südpazifiks primär im Kontext RELIGION auf (z. B. *Esel, Schlange, Kamel, Löwe*). Die sich hyperonymisch einheitlich verhaltende Wortgruppe fällt also thematisch in zwei verschiedene Felder.

Ausgehend von diesen Überlegungen haben wir versucht, ein thematisches Klassifikationsschema induktiv aus Lehnwortdaten zu gewinnen. Induktiv soll hier nicht heißen, dass das Wissen über bestehende Systeme (wie etwa LTML, Dornseiff 1934, 2004, Hallig/Wartburg 1952, Archer et al. 2004) vollständig aus-

5 Vgl. zu ähnlichen Überlegungen zur Rolle von Kookkurrenzen bei der Zuordnung von lexikalischen Einheiten zu Feldern/Sachgruppen auch Wiegand (2004) und Quasthoff (2004).

6 Die Formulierung „mit den weitaus meisten“ beinhaltet, dass es möglich ist, eine Wortbedeutung mehr als nur einem thematischen Feld zuzuordnen.

7 Diese Konzeptionalisierung und Operationalisierung des Begriffs „thematisches Feld“ ist eher grundsätzlicher Natur. Die Operationalisierungen werden im Einzelfall im Rahmen der hier anvisierten lexikographischen Annotation nicht praktikierbar sein.

geblendet wird, sondern dass thematische Felder auf der Basis systematischen Analysierens von Lehnwortbedeutungen etabliert werden. Dazu wurde eine Liste von 450 Lehnwortbedeutungen zusammengestellt, die sich aus jeweils einer Zufallsauswahl von 150 Einheiten aus *W-Slv*, *W-Frz* und *W-Pzf* zusammensetzte. Um die thematische Annotation nicht zu erschweren und bei onomasiologischen Abfragen nicht zu kleine Treffermengen zu bieten, sollte die entstehende Klassifikation nicht zu feingliedrig sein. Angestrebt wurde eine maximal zweistufige Klassifikation, die auf der obersten Stufe nicht mehr als einige Dutzend Klassen enthält. Aus der sukzessiven Bearbeitung der Lehnwortliste entwickelten wir eine Klassifikation, die als alphabetisch geordnete Themenliste so beginnt:

- | | |
|---------------------|-------------------------|
| 01: BAUWESEN | 09: HAUSHALT |
| 02: BERGBAU | 10: HIMMEL |
| 03: CHARAKTER | 11: JAGD & FISCHFANG |
| 04: ESSEN & TRINKEN | 12: KLEIDUNG & SCHMUCK |
| 05: FAMILIE | 13: KOMMUNIKATION |
| 06: GELD | 14: KRANKHEIT & HEILUNG |
| 07: GESELLSCHAFT | 15: KULTURRAUM |
| 08: HANDWERK | 16: ... |

Einige der Themen erlauben für einen größeren Teil der ihnen zugeordneten lexikalischen Einheiten eine Klassifikation auf zweiter Stufe, z. B.

- LANDWIRTSCHAFT
- LANDWIRTSCHAFT > ACKERBAU & OBST- UND GEMÜSEANBAU
- LANDWIRTSCHAFT > HAUSTIERE & VIEHZUCHT
- LANDWIRTSCHAFT > WEINBAU
- WISSENSCHAFT
- WISSENSCHAFT > BIOLOGIE
- WISSENSCHAFT > CHEMIE
- WISSENSCHAFT > PHYSIK
- WISSENSCHAFT > PSYCHOLOGIE
- WISSENSCHAFT > ...

Die Themenbezeichnungen wurden dabei jeweils mit Annotationshinweisen versehen, die den Umfang des Themas illustrieren, z. B.:

- BAUWESEN** Bezeichnungen für Gebäude, Gebäudeteile, Räumlichkeiten, Baumaterialien, Handlungen beim Errichten von Gebäuden

HAUSHALT Bezeichnungen für Gegenstände und Handlungen, die primär dem privaten Haushalt zuzurechnen sind (einschließlich Kochen und Küche, aber ohne KLEIDUNG & SCHMUCK und ohne ESSEN & TRINKEN)

KRANKHEIT & HEILUNG Bezeichnungen für Krankheiten und Ausdrücke für körperliche Unwohlseinsempfindungen; Bezeichnungen für Heilmittel, Bepflanzungen im Gesundheitsbereich, Handlungen des Heilens

Letztlich konnten über diese Vorgehensweise etwa 90 % der Lehnwortbedeutungen einem oder mehreren Themen zugeordnet werden.⁸ Hier ein Auszug aus der Lehnwortliste mit den entsprechenden Zuordnungen:

Wortbedeutung	Zielsprache	Lehnwort	Quelle	Thematisches Feld
...
Reis	Tok Pisin	<i>rais</i>	W-Pzf	(i) Landwirtschaft > Ackerbau, (ii) Essen & Trinken
Relativität(stheorie)	Französisch	<i>relativité</i>	W-Frz	Wissenschaft > Physik
Rollmops	Französisch	<i>rollmops</i>	W-Frz	Essen & Trinken
Rotpunkt	Französisch	<i>rot-punkt</i>	W-Frz	Spiel & Sport
Sägemehl	Tok Pisin	<i>sikmel</i>	W-Pzf	Handwerk
Satan	Nauruisch	<i>Satan</i>	W-Pzf	Religion
Schaf	Nauruisch	<i>Schaf</i>	W-Pzf	(i) Landwirtschaft > Haustiere-& Viehzucht, (ii) Religion
Schaf, graue Geiß	Slovenisch	<i>birka</i>	W-Slv	Landwirtschaft > Haustiere-& Viehzucht
scheiden	Französisch	<i>scheider</i>	W-Frz	Bergbau
...

Bei den verbleibenden ca. 45 Lemmata ergaben sich Zuordnungsprobleme verschiedener Art:

- Die lexikographische Bedeutungsbeschreibung des Lehnworts war unklar, z. B. Slov. *ahker* 'die Warte' < Kärntnerisch *árker* (Neuhochdeutsch *Erker*).⁹

8 Eine Überprüfung der Annotatorenübereinstimmung bei der Kodierung steht noch aus.

9 Die im vorliegenden Fall vermutlich einschlägige Bedeutung „Wachturm“ o. ä. ist als Wortbedeutung von *Warte* im heutigen Deutsch außerhalb von Toponymen kaum mehr gebräuchlich. Dass die Lesart „erhöhter Standpunkt“ nicht gemeint war, lässt sich

- Das Lehnwort war nicht hinreichend disambiguiert, z. B.: Yapesisch *silinder* < Dt. *Zylinder*. Es ist unklar, ob es sich um das Kleidungsstück oder das Maschinenteil handelt.
- Es ließ sich kein geeignetes Feld konstruieren, z. B. Slov. *ajntajlung* < Dt. *Einteilung*.

Auch wenn mit dem induktiven Klassifikationsversuch insgesamt ein befriedigendes Ergebnis erzielt wurde, das eine für onomasiologische Abfragen vermutlich interessante Klassenbildung ermöglicht, sind natürlich auch erhebliche Schwächen eines solchen Verfahrens zu konstatieren:

- i) Die Setzung von Themen ist zwar durch die Daten geleitet, aber nur schwach kriteriengesteuert und nicht operationalisiert.
- ii) Die Klassifikation ist spezifisch für unseren Lehnwortschatz und nicht ohne Weiteres auf andere Klassifikationen abbildbar, was wiederum vergleichende Studien (etwa mit der World Loanword Database) einschränkt.
- iii) Das Inventar an Klassen erfordert vermutlich mit der Erweiterung des Lehnwortbestands auch Erweiterungen oder Änderungen des Themeninventars. Damit werden wiederum möglicherweise Reklassifizierungen des lexikalischen Bestands erforderlich.

Überprüft wurde die Lehnwortliste auch hinsichtlich hyperonymischer Relationen. Insgesamt war der Bestand an lexikalischen Einheiten, die aus Sicht einer onomasiologischen Suche sinnvollerweise einem Hyperonym zugeordnet werden konnten, geringer als bei der thematischen Zuordnung. Insbesondere die Hyperonyme *Fahrzeug*, *Gebäude*, *Zahlungsmittel*, *Kleidungsstück*, *Lebensmittel*, *Maßeinheit*, *Möbelstück*, *Person hinsichtlich ihres Berufs*, *Pflanze*, *Tier*, *Waffe* und *Werkzeug* subsumierten jeweils eine größere Anzahl an Lehnwortbedeutungen. Grundsätzlich könnten auch meronymische Relationen für eine onomasiologische Abfrage von Interesse sein. In dem untersuchten Lehnwortbestand fanden sich vereinzelt solche Relationen, etwa bei den Lehnwörtern zu den Etyma *Stachelbeere* als Meronym von *Stachelbeerstrauch* > *Pflanze* oder *Erker* als Meronym zu *Gebäude*. Wir haben diese Option hier nicht weiter verfolgt.

dem Wörterbuchartikel, der die alleinige Grundlage für die ontologische Zuordnung darstellt, nicht explizit entnehmen, sondern erschließt sich dem philologisch geschulten Leser allenfalls aufgrund der weiteren etymologischen Erläuterungen im Artikel.

3. Onomasiologische Zugriffsstruktur im *Lehnwortportal Deutsch*

Im Folgenden soll geklärt werden, wie sich ein onomasiologisches System sinnvoll in die Datenstrukturen des Lehnwortportals integrieren und für onomasiologische Suchen nutzbar machen lässt. Der einfacheren Exposition halber wird im Folgenden lediglich das Problem thematischer Kodierungen von Wortbedeutungen berücksichtigt; hyperonymische Information lässt sich völlig analog modellieren.

3.1. Zur graphenbasierten Datenorganisation im *Lehnwortportal Deutsch*

Mit Ausnahme der Artikel des „umgekehrten Lehnwörterbuchs“ ist jeder Lehnwörterbucheintrag im Lehnwortportal Deutsch datenbankseitig zunächst als eigenständiges XML-Dokument repräsentiert, aus dem, außer bei imagedigitalisierten Werken, mit Standardtechnologien wie XSLT auch die gewöhnliche HTML-basierte Online-Ansicht des jeweiligen Artikels generiert wird. Nach einem für jedes Wörterbuch separat zu definierenden Verfahren (vgl. Meyer 2014b) wird zusätzlich aus jedem der XML-Dokumente automatisch die Menge der in ihm repräsentierten *Relationen* zwischen gebuchten Wortformen (Lehnwörter und Etyma sowie davon abgeleitete Lexeme, insbesondere Derivate) extrahiert. Wichtige Typen von Relationen zwischen Lexemen sind etwa „y ist Lehnwort zum Etymon x“, „y ist Derivat zu x“, „y ist Kompositum mit Kompositionsglied x“, „y ist (phonologische, diasystematische, ...) Variante zu x“. Hinzu kommen Beziehungen des Typs „Etymon y ist dem Portal-Metalemma x zugeordnet“, die – mitsamt den Metalemmata selbst – beim Prozess der Integration der Lehnwörterbücher in das Portal mit eigens entwickelten Softwarewerkzeugen manuell definiert werden müssen und zugleich als Datengrundlage für das „umgekehrte Lehnwörterbuch“ dienen. So entsteht ein komplexes Netz von Beziehungen zwischen Wörtern, das formal als gerichteter Graph beschreibbar ist und im Portal im Rahmen einfacher Visualisierungen auch dem Benutzer anschaulich vermittelt wird. Nachstehende Abb. 1 zeigt den vom Metalemma *Waage* ausgehenden Teilgraphen, wie er unter <http://lwp.ids-mannheim.de/art/meta/lemma/Waage> angezeigt wird, und deutet zusätzlich die Quellen der visualisierten Informationen an.

Die Graphenrepräsentation wird datenbankseitig separat verwaltet, derzeit in einer relationalen Datenbank und im parallelen Testbetrieb auch in einer dedizierten Graphendatenbank. Diese Datenbanken werden bei der Erstellung der Artikel des „umgekehrten Lehnwörterbuchs“ verwendet, insbesondere aber auch für die sehr komplexen, z. T. wörterbuchübergreifenden Rechercheoptionen des Portals (vgl. Meyer 2013a, 2013b für weitere Details).

3.2. Integration eines onomasiologischen Systems in die Datenmodellierung

Jedem als Knoten im Graphen repräsentierten Wort/Lexem lassen sich Eigenschaften (Attribute) zuordnen, etwa zur orthografischen Form sowie zu grammatischen Eigenschaften und zur diastematischen Einordnung der betreffenden Form. Im Lehnwortportal Deutsch wird ein „Steckbrief“ der mit einem Knoten assoziierten Eigenschaften in der bereits erwähnten Graphenvisualisierung angezeigt, wenn man den Mauszeiger über dem betreffenden Knoten ruhen lässt. Insbesondere wird derzeit eine Liste aller zu einem Wort angegebenen lexikographischen Bedeutungsdefinitionen als Knotenattribut geführt und ermöglicht es Benutzern, in den fortgeschrittenen Suchfunktionen auch nach Zeichenketten in Bedeutungsangaben zu suchen. Da vor allem bei Komposita und produktiv gebildeten Derivaten in der Zielsprache in Wörterbüchern gerne auf eine explizite Bedeutungsangabe verzichtet wird, ist dieses Attribut in nicht wenigen Fällen jedoch leer bzw. nicht vorhanden.

Um die einzelnen Wortbedeutungen thematischen Feldern zuzuordnen, könnte man die Bedeutungsangaben selber als Listen komplex strukturierter Datenobjekte – statt einfach als Listen von Zeichenketten – modellieren. Ein Attribut eines solchen Datenobjektes wäre dann eine nicht leere Liste von Codes für die zugeordneten thematischen Felder. In der derzeit im Portal verwendeten relationalen Implementierung des Graphen ist ein solches Vorgehen technisch ohne weiteres möglich. Der bereits vorhandenen 1:n-Relation zwischen einem Knoten und seinen Bedeutungen wäre eine m:n-Relation zwischen diesen Bedeutungen und der Liste der thematischen Felder hinzuzufügen. In üblichen Graphendatenbanken stößt man mit diesem Vorgehen allerdings an eine Grenze, da beliebig komplex strukturierte Knotenattribute im Allgemeinen nicht möglich und auch technisch nicht sinnvoll sind. Der naheliegende Ausweg besteht darin, die Bedeutungen eines Wortes ebenfalls graphentheoretisch zu modellieren, also einen neuen Knotentyp einzuführen, der kein Wort bzw. Lexem, sondern eine einzelne Wortbedeutung repräsentiert und als Attribute zum einen die im jeweiligen Wörterbuch angegebene Bedeutungsdefinition, sofern vorhanden, und zum anderen

eine Liste der zugeordneten thematischen Felder aufweist. Von einem Wortknoten können dann eine oder mehrere gerichtete Kanten zu den zugehörigen Wortbedeutungsknoten führen. Ein technisch-konzeptueller Vorteil dieser Modellierung ist die einfachere und besser indizierbare Abfragbarkeit; es gibt aber auch lexikographische Vorteile einer ‚Reifizierung‘ von Wortbedeutungen als Knoten, die es insbesondere ermöglicht, beliebige Relationen zwischen Wortbedeutungen zu modellieren. Wenn beispielsweise bei einem Derivat, das die Nehmersprache nach einem produktiven (und dem intendierten menschlichen Adressaten vertrauten) Wortbildungsmuster von einem Lehnwort gebildet hat, die Bedeutungsangabe fehlt, kann in der graphenbasierten Modellierung ein „Platzhalter“-Bedeutungsknoten angesetzt werden, der durch Kanten eines speziellen, verweisenden Typs mit den im Wörterbuch definierten Bedeutungen des Simplex verbunden ist. Interessanter sind Beziehungen zwischen den Bedeutungen von Lehnwörtern in Entlehnungsketten. So werden in einem laufenden Projekt zu polnisch vermittelten Germanismen in den ostslavischen Sprachen, deren lexikographische Erfassung im Lehnwortportal publiziert werden soll (Meyer 2014c), systematisch die Beziehungen von Wortbedeutungen der ostslavischen Lehnwörter zu den Bedeutungen korrespondierender Lehnwörter laut dem bereits im Portal republizierten *W-Pol* erfasst; es ist daher denkbar, dass solche Beziehungen zumindest für diesen Teilbereich des Portals in die Graphendatenbank aufgenommen werden und für Benutzer abfragbar sind.

3.3. Onomasiologische Suchfunktionen im Lehnwortportal

Schon jetzt sind im *Lehnwortportal Deutsch* wörterbuchübergreifende Suchen nach Zeichenketten in Bedeutungsangaben möglich, die zudem mit Booleschen Operatoren miteinander verknüpfbar sind. In der formularbasierten erweiterten Suche¹⁰ kann so beispielsweise nach Wörtern gesucht werden, bei denen in der Definition einer und derselben Wortbedeutung sowohl die Zeichenkette „Gewicht“ als auch die Zeichenkette „Last“ vorkommt; alternativ ist aber auch eine Suche nach Wörtern möglich, bei denen sowohl das Wort „Gewicht“ als auch das Wort „Wichtigkeit“ in je mindestens einer Lesart, aber nicht notwendigerweise in derselben, vorkommen. In einem um thematische Felder erweiterten Suchformular könnte man dann zusätzlich eine Auswahlliste der in der thematischen Klassifikation verwendeten Felder anbieten, so dass der Nutzer der Suchfunktion Wortbedeutungen sowohl unter Angabe einer in der Definition enthaltenen Zeichenkette als auch anhand zugeordneter thematischer Felder spezifizieren kann.

10 Vgl. <<http://lwp.ids-mannheim.de/search/meta>> [aufgerufen am 04.07.2015].

Für wissenschaftliche Zwecke besonders interessant sind die bereits jetzt möglichen Abfragen mit einer eigens für das Lehnwortportal konzipierten, deklarativen Abfragesprache¹¹ die es ermöglicht, explizit nach beliebig komplexen Konstellationen im gerichteten Graphen zu suchen.¹² Werden die Wortbedeutungen wie beschrieben als Knoten kodiert, lässt sich die Abfragesprache auf naheliegende Weise so erweitern, dass auch die thematische Klassifikation von Bedeutungen in Suchen einbezogen werden kann. Ein hypothetisches¹³ Beispiel:

```
/* deklariere und benenne Knoten (Wörter) */
```

```
suche etymon ety.
```

```
suche lehnwort lw.
```

```
suche bedeutung lw_bed.
```

```
suche lehnwort deriv.
```

```
suche bedeutung deriv_bed.
```

```
/* lege Eigenschaften der Knoten (Wörter) fest */
```

```
ety ist substantiv.
```

```
lw_bed gehoert zum feld 'Bewertung & Klassifikation'.
```

```
deriv ist adjektiv.
```

```
deriv_bed enthaelt 'wichtig'.
```

```
/* lege Eigenschaften der Kanten (Relationen) fest */
```

```
lw ist lehnwort zu ety.
```

```
lw_bed ist bedeutung von lw.
```

```
deriv ist derivat zu lw.
```

```
deriv_bed ist bedeutung von deriv.
```

Mit der gezeigten Abfrage werden Teilgraphen mit folgender Konstellation gesucht: Ein Lehnwort *lw* zu einem deutschen substantivischen Etymon *ety* bildet nehmersprachlich ein adjektivisches Derivat *deriv*. Mindestens eine Wortbedeutung *lw_bed* von *lw* ist dem thematischen Feld ‚Bewertung & Klassifikation‘ zugeordnet; die Definition mindestens einer Wortbedeutung *deriv_bed* von *deriv* enthält die Zeichenkette *wichtig*. Einer solchen Suchanfrage entspricht als Suchergebnis allgemein eine Liste von passenden Knoten-*n*-Tupeln. Im vorliegenden Beispiel könnte ein einzelnes konkretes Suchergebnis etwa so dargestellt werden:

11 Vgl. <<http://lwp.ids-mannheim.de/search/prof>> [aufgerufen am 04.07.2015].

12 Vgl. hierzu die allgemeinen Überlegungen in Meyer (2013a) sowie die Dokumentation des Abfragesystems unter <<http://lwp.ids-mannheim.de/doc/meta/graphsearch>> [aufgerufen am 04.07.2015].

13 Lässt man im Beispiel die drei Bedingungen zu Bedeutungsknoten fort, entsteht eine bereits jetzt im Portal zulässige Suchanfrage.

- *ety*: Etymon *Waage* (Subst., f.) im Artikel → *waga* (W-Pol)
- *lw*: Lehnwort *waga* (Subst., f.) im Artikel → *waga* (W-Pol)
- *lw_bed*: Bedeutung 'Wichtigkeit, Bedeutung von etw.'; thematische Felder: BEWERTUNG & KLASSIFIKATION, ...; hyperonymische Relationen: ... im Artikel → *waga* (W-Pol)
- *deriv*: Derivat *ważny* (Adj.) im Artikel → *waga* (W-Pol)
- *deriv_bed*: Bedeutung 'wichtig; gewichtig'; thematische Felder: BEWERTUNG & KLASSIFIKATION, ...; hyperonymische Relationen: ... im Artikel → *waga* (W-Pol)

4. Literatur¹⁴

- ARCHER, DAWN; RAYSON, PAUL; PIAO, SCOTT; MCENERY, TONY (2004): Comparing the UCREL Semantic Annotation Scheme with Lexicographical Taxonomies. In: Williams, Geoffrey; Vessier, Sandra (eds.): *Proceedings of the Eleventh EURALEX International Congress. EURALEX 2004, Lorient, France. July 6–10, 2004. Volume III*. Lorient Cedex: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, 817–827.
- BUSSE, DIETRICH; TEUBERT, WOLFGANG (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich; Herrmanns, Fritz; Teubert, Wolfgang (eds.): *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Oplanden: Westdeutscher Verlag, 10–28.
- DORNSEIFF, FRANZ (1934): *Der deutsche Wortschatz nach Sachgruppen*. Berlin: de Gruyter.
- DORNSEIFF, FRANZ (2004): *Der deutsche Wortschatz nach Sachgruppen*. 8., völlig neu bearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Berlin: de Gruyter.
- ENGELBERG, STEFAN (2010): An inverted loanword dictionary of German loanwords in the languages of the South Pacific. In: Dykstra, Anne; Schoonheim, Tanneke (eds.): *Proceedings of the XIV EURALEX International Congress (Leeuwarden, 6–10 July 2010)*. Ljouwert (Leeuwarden): Fryske Akademy, 639–647.
- FELLBAUM, CHRISTIANE (ed.) (1998): *WordNet: An Electronic Lexical Database*. Cambridge: MA: MIT Press.
- HALLIG, RUDOLF; VON WARTBURG, WALTHER (1952): *Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*. Berlin: Akademie-Verlag.

14 Alle im Folgenden angegebenen URL's wurden am 29.5.2015 aufgerufen.

- HAMP, BIRGIT; FELDWEIG, HELMUT (1997): GermaNet – a Lexical-Semantic Net for German. In: *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. ACL/EACL-97 Workshop Proceedings. July 12th 1997. Madrid*. Somerset: Association for Computational Linguistics, 9–15.
- HASPELMATH, MARTIN; URI, TADMOR (eds.) (2009a): *World Loanword Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wold.clld.org>>
- HASPELMATH, MARTIN; TADMOR, URI (2009b): The Loanword Typology project and the World Loanword Database. In: Martin Haspelmath; Tadmor, Uri (eds.): *Loanwords in the world's Languages. A Comparative Handbook*. Berlin: de Gruyter: 1–34.
- HENRICH, VERENA; HINRICHS, ERHARD (2010): GernEiT – The GermaNet Editing Tool. In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC10)*. Valletta, Malta. May 17–23, 2010. Valletta, Malta: European Language Resources distribution agency, 2228–2235.
- IZQUIERDO BEVIÁ, RUBÉN; SUÁREZ CUETO, ARMANDO; RIGAU CLARAMUNT, GERMAN (2007): Exploring the Automatic Selection of Basic Level Concepts. In: *International Conference. Recent Advances in Natural Language Processing. Proceedings*. Borovets, Bulgaria. Shoumen: INCOMA Ltd., 298–302.
- LIN, SHU-YEN [林淑晏] (2010): *A Computational Study of the Basic Level Nouns in English*. Unpublished doctoral dissertation, National Taiwan Normal University. <http://ir.lib.ntnu.edu.tw/retrieve/51190/metadata_02_02_s_05_0155.pdf>.
- MEYER, PETER; ENGELBERG, STEFAN (2011): Ein umgekehrtes Lehnwörterbuch als Internetportal und elektronische Ressource: Lexikographische und technische Grundlagen. In: Hedeland, Hanna; Schmidt, Thomas; Wörner, Kai (eds.): *Multilingual Resources and Multilingual Applications*. Hamburg: Universität Hamburg, 169–174.
- MEYER, PETER (2013a): Advanced graph-based searches in an Internet dictionary portal. In: Kosem, Iztok et al. (eds.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn, 488–502.
- MEYER, PETER (2013b): Ein Internetportal für deutsche Lehnwörter in slavischen Sprachen. Zugriffsstrukturen und Datenrepräsentation. In: Kempgen, Sebastian et al. (eds.): *Deutsche Beiträge zum 15. Internationalen Slavistenkongress, Minsk 2013* (= Die Welt der Slaven. Sammelbände 50). München: Sagner, 233–242.
- MEYER, PETER (2014a): Graph-Based Representation of Borrowing Chains in a Web Portal for Loanword Dictionaries. In: Abel, Andrea; Vettori, Chiara; Ralli, Natascia (eds.): *Proceedings of the XVI EURALEX International Congress:*

- The User in Focus. 15–19 July 2014, Bolzano/Bozen.* Bolzano/Bozen: EURAC research, 1135–1144.
- MEYER, PETER (2014b): Von XML zum DAG: Der lexikographische Prozess bei der Erstellung eines graphenbasierten Wörterbuchportals. In: Domínguez Vázquez, María José; Mollica, Fabio; Nied Curcio, Martina (eds.): *Zweisprachige Lexikographie zwischen Translation und Didaktik* (= *Lexicographica*, Series Maior 147). Berlin: de Gruyter, 303–321.
- MEYER, PETER (2014c): Entlehnungsketten in einem Internetportal für Lehnwörterbücher. IT-Infrastruktur und computerlexikographischer Prozess in einem Projekt zu polnisch vermittelten Germanismen im Ostslavischen. In: Mann, Michael (ed.): *Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen* (= *Germanistische Linguistik* 223/224). Hildesheim u. a.: Olms, 97–132.
- NILES, IAN; PEASE, ADAM (2001): Towards a standard upper ontology. In: Welty, Chris; Smith, Barry (eds.): *2nd International Conference on Formal Ontology in Information Systems. FOIS 2001, Ogunquit, Maine, USA. October 17–19, 2001. Proceedings*. New York, NY: ACM Press, 17–19.
- NILES, IAN; PEASE, ADAM (2003): Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Arabnia, Hamid R. (ed.): *Proceedings of the 2003 International Conference on Information and Knowledge Engineering IKE'03. June 23–26, 2003. Las Vegas, Nevada, USA. Volume II*. Athens, GA: CSREA Press, 412–416.
- QUASTHOFF, UWE (2004): Methodologische Einführung. In: Dornseiff, Franz: *Der deutsche Wortschatz nach Sachgruppen*. 8., völlig neubearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Berlin u. a.: de Gruyter.
- ROSCH, ELEANOR (1977): Human Categorisation. In: Warren, Neil (ed.): *Studies in Cross-Cultural Psychology*, Vol. I. London: Academic Press, 1–49.
- SARCHER, WALBURGA (2001): *Das deutsche Lehngut im Französischen als Zeugnis für den Wissenstransfer im 20. Jahrhundert*. Hamburg: Kovač.
- STRIEDTER-TEMPS, HILDEGARD (1963): *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.
- VINCENZ, ANDRÉ DE; HENTSCHEL, GERD; BRÜGGEMANN, MARK (2010): *Wörterbuch der deutsche Lehnwörter in der polnischen Schrift- und Standardsprache. Von den Anfängen des polnischen Schrifttums bis in die Mitte des 20. Jahrhunderts* (= *Studia Slavica Oldenburgensia* 20). Oldenburg: BIS-Verlag. <<http://diglib.bis.uni-oldenburg.de/bis-verlag/wdlp>>.

- VOSSEN, PIEK ET AL. (1998): *The EuroWordNet Base Concepts and Top Ontology. Deliverable D017, D034, D036, WP5, EuroWordNet (LE2-4003)*. University of Amsterdam.
- WIEGAND, HERBERT ERNST (2004): Lexikographisch-historische Einführung. In: Dornseiff, Franz (2004): *Der deutsche Wortschatz nach Sachgruppen*. 8., völlig neubearbeitete und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Berlin: de Gruyter: 9–91.